Polya's Urn

Math 365 Tanya Leise

April 28, 2022

The objective of this lab is to explore an example of a martingale called "Polya's urn."

Submit your completed lab to Gradescope.

Description of Polya's Urn

Suppose you have an urn containing one red ball and one green ball. You draw one at random; if the ball is red, put it back in the urn with an additional red ball (so urn now has 2 red and 1 green), otherwise put it back and add a green ball (so urn now has 1 red and 2 green). Repeat this procedure and let the random variable X_n be the number of red balls in the urn after n draws. Note that $X_0 = 1$, and after n draws there will be $n + 2$ total balls in the urn with $1 \leq X_n \leq n + 1$.

The random variables X_0, X_1, X_2, ... form a Markov chain with the following transition probabilities for $1 \leq k \leq n + 1$:

$$\mathbb{P}\{X_{n+1} = k \,|\, X_n = k\} = 1 - \frac{k}{n+2}$$

$$\mathbb{P}\{X_{n+1} = k+1 \,|\, X_n = k\} = \frac{k}{n+2}$$

Let $M_n = \frac{1}{n+2}X_n$, the proportion of red balls in the urn after n draws. To prove that M_n is a martingale with respect to X_n, it's sufficient to show that $E(M_{n+1} \,|\, X_n) = M_n$. Check the calculations below to make sure they make sense to you:

$$E(X_{n+1} \,|\, X_n = k) = k\mathbb{P}\{X_{n+1} = k \,|\, X_n = k\} + (k+1)\mathbb{P}\{X_{n+1} = k+1 \,|\, X_n = k\} = \frac{n+3}{n+2}k,$$

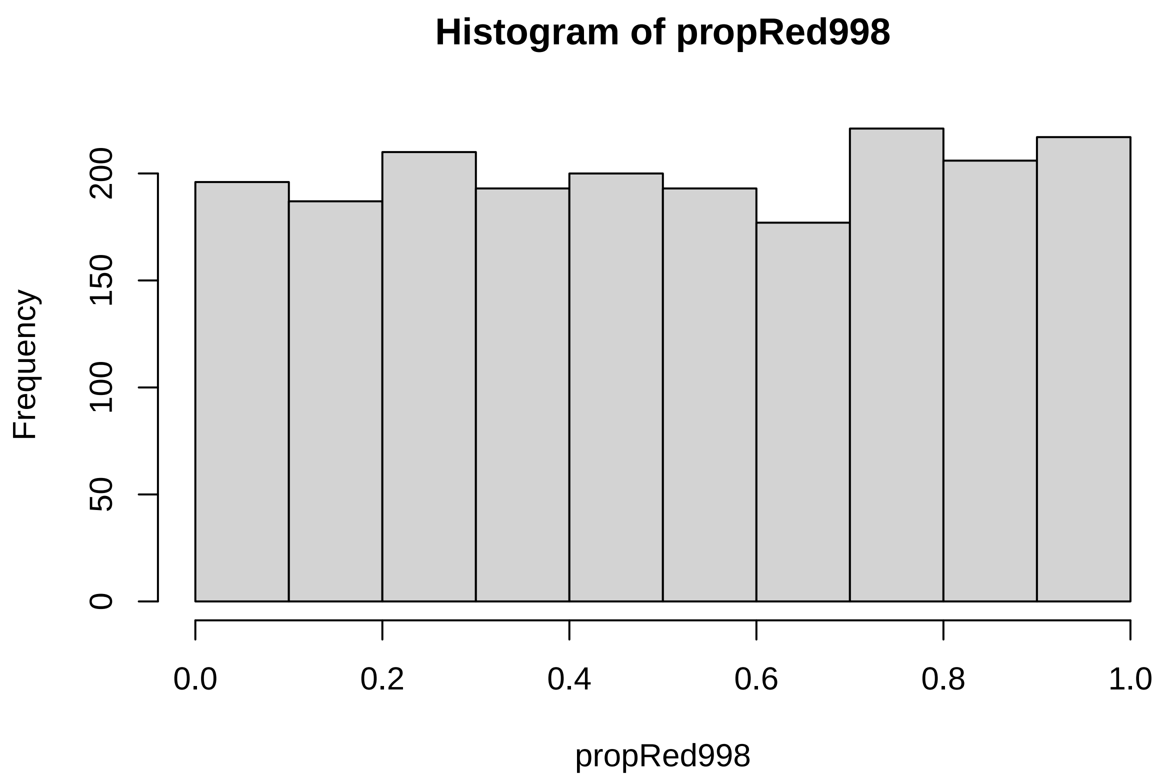
Using the definition, $M_{n+1} = \frac{1}{n+3}X_{n+1}$, so

$$E(M_{n+1} \,|\, X_n) = \frac{1}{n+3}E(X_{n+1} \,|\, X_n) = \frac{1}{n+3} \cdot \frac{n+3}{n+2}X_n = \frac{1}{n+2}X_n = M_n.$$

Simulation of Polya's urn

The following R code simulates 998 draws, until the urn contains 1000 balls. The simulation is run 2000 times, so you can examine the distribution of M_{998} at the end:

```r
nStop <- 998
nSim <- 2000
propRed998 <- matrix(0,nSim,1)
for (n in 1:nSim) {
  nRed <- 1
  nGreen <- 1
  for (k in 1:nStop) {
    drawnball <- sample(0:1,size=1,prob=c(nRed,nGreen)/(nRed+nGreen))
    nRed <- nRed+(1-drawnball) # drawnball=0 if red, 1 if green
    nGreen <- nGreen+drawnball
    }
  propRed998[n] <- nRed/(nRed+nGreen)
  }
hist(propRed998)
```



Exercise 1

Conjecture what the distribution of M_n is in general, based on the simulation results.

Answer: The propRed possess a uniform distribution because we can see that differernt values of propRed have similar frequencies.

Exercise 2

Prove using induction on n that $\mathbb{P}\{M_n = \frac{k}{n+2}\} = \frac{1}{n+1}$ for $k = 1, 2, \ldots, n+1$. That is, the distribution is uniform. Hint: as we often do with Markov chain calculations, condition on the previous step, considering how to get to k red balls in the urn in going from the nth to $(n+1)$th step.

Prob 2

$$M_n = \frac{X_n}{n+2}$$

$$\mathbb{P}\left(M_n = \frac{k}{n+2}\right) = \frac{1}{n+1} \quad \text{for } k = 1, 2, \ldots, n+1$$

$$\left(\mathbb{P}(X_n = k) = \frac{1}{n+1}\right)$$

How to get k red balls in the urn in going from n to $(n+1)^{th}$ step.

$$\Rightarrow P\left(M_1 = \frac{k}{3}\right) = \frac{1}{2}$$

$\boxed{1R1B}$ $\Rightarrow$ for $k=1$: $1/2$ prob

$\Rightarrow$ for $k=2$: $1/2$ prob

$$\boxed{\text{Assuming } P\left(M_{n-1} = \frac{k}{n+1}\right) = \frac{1}{n} \ \forall\, k \in \{1, 2, \ldots, n\}}$$

$$P\left(M_n = \frac{k}{n+2}\right) = \left(\frac{(n+1)-k}{(n+1)}\right) \underbrace{P\left(M_{n-1} = \frac{k}{n+1}\right)}_{1/n} + \left(\frac{k-1}{n+1}\right) \underbrace{P\left(M_{n-1} = \frac{k-1}{n+1}\right)}_{1/n}$$

assumed by induction

$$= \left(1 - \frac{k}{n+1}\right)\left(\frac{1}{n}\right) + \left(\frac{k-1}{n+1}\right)\left(\frac{1}{n}\right)$$

$$= \frac{1}{n}\left(1 - \frac{1}{n+1}\right) = \boxed{\frac{1}{n+1}}$$

Hence Proved

Figure 1: Ex 2

Another simulation of Polya's urn

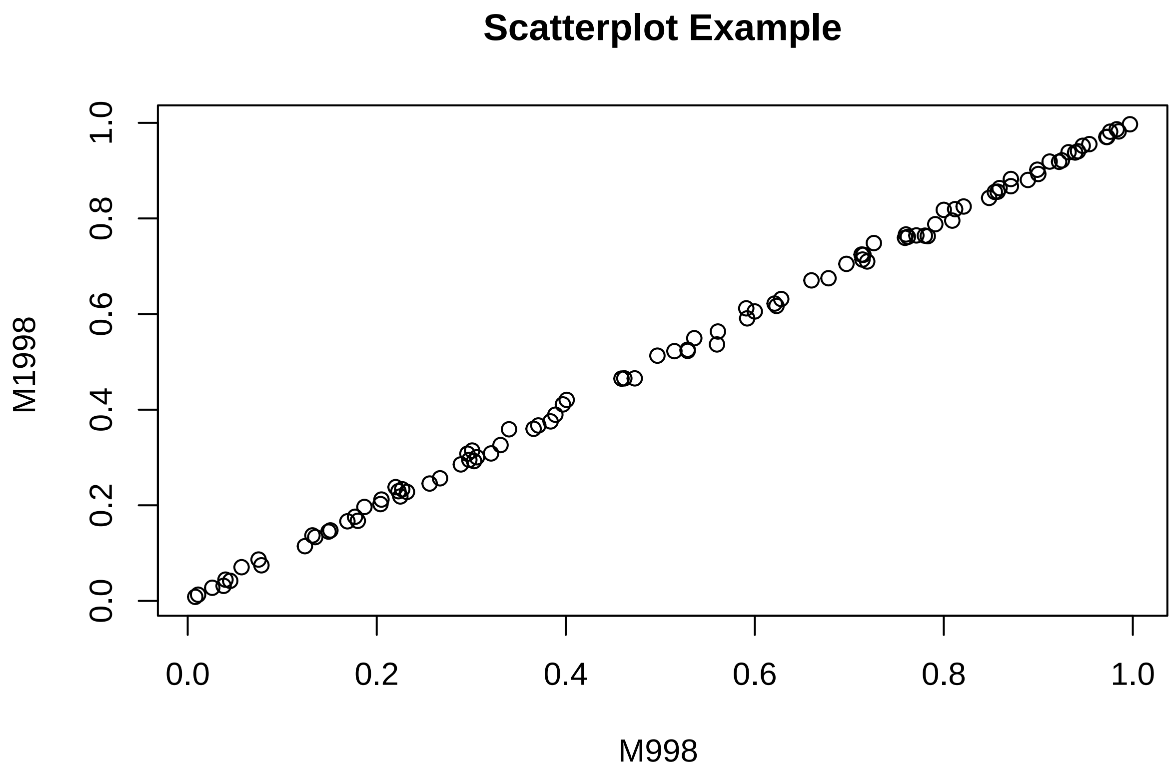
Exercise 3

Revise the simulation code to run 100 simulations that each do 1,998 draws (so total of 2,000 balls when done). Store the proportion of red balls after 998 draws and also after 1,998 draws. How do these numbers compare for each simulation? Create a scatter plot of M_{1998} versus M_{998} from the 100 simulations, e.g., plot(M998,M1998). Observe how the values at the two timepoints are highly correlated: once the set is large enough, the proportion of red tends to stay about the same for a long time.

Solution: We can see that the two proportions are highly correlated and the proportion of red tends to stay about the same for a long time.

```
nStop <- 998
nSim <- 100
propRed1998 <- matrix(0,nSim,1)
propRed998 <- matrix(0,nSim,1)
for (n in 1:nSim) {
  nRed <- 1
  nGreen <- 1
  for (k in 1:nStop) {
    drawnball <- sample(0:1,size=1,prob=c(nRed,nGreen)/(nRed+nGreen))
    nRed <- nRed+(1-drawnball) # drawnball=0 if red, 1 if green
    nGreen <- nGreen+drawnball
  }
  propRed998[n] <- nRed/(nRed+nGreen)
  for (k in 999:1998) {
    drawnball <- sample(0:1,size=1,prob=c(nRed,nGreen)/(nRed+nGreen))
    nRed <- nRed+(1-drawnball) # drawnball=0 if red, 1 if green
    nGreen <- nGreen+drawnball
    }
  propRed1998[n] <- nRed/(nRed+nGreen)
  }
#hist(propRed1998)

plot(propRed998, propRed1998, main="Scatterplot Example",
   xlab="M998 ", ylab="M1998 ")
```


Scatterplot Example
M1998
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
M998